Tiedetta Vai Tarinaa podcast jakso 2

Uhkaako tekoäly ihmiskuntaa kuten scifissä? Koneoppimisen tutkija vastaa

Suvi: Oletko sinä Arno jo rakentanut oman bunkkerin tätä tekoälyn vallankaappausta varten?

Arno: En ole rakentanut, enkä ole valmistautunutkaan tähän näin. Ja tietysti mitä enemmän näitä asioita pohtii, niin sen ajankohtaisemmaksi tämä tuntuu muuttuvan.
Suvi: Tervetuloa seuraamaan Tiedettä vai tarinaa -podcastia, jossa tarkastelemme scifitarinoita tieteellisestä näkökulmasta. Mun nimi on Suvi Rantonen. Olen Aalto-yliopiston Sähkötekniikan korkeakoulun alumni, teknologiayritys Vaisalan prosessijohtaja ja innokas scifi-harrastaja. Tänään kanssani näitä tarinoita tutkiskelee Aalto-yliopiston huippuasiantuntija koneoppimisen saralta, viimeisimmäksi vuoden tekoälytutkijana palkittu apulaisprofessori Arno Solin. Lämpimästi tervetuloa!

Arno: Kiitoksia!

Tekoälyjuontaja: Yllätys! Mukana olen myös minä, tekoälyjuontaja Suvi. Minut on luotu tosimaailman Suvin pohjalta, mutta digitaalisena olentona minulla ei ole biologian tuomia rajoitteita ja olen yli-inhimillisen älykäs. Osaan sanoa olympia eikä olumpia, floridan broileri, psykiatrikologi, meteorologideodorantti ja express...

Suvi: Ihmisten rakentamien koneiden ennalta-arvaamaton käyttäytyminen on ollut suosittu teema populaarikulttuurissa aina Frankensteinin hirviöstä lähtien jo yli sadan vuoden ajan. Koneiden ylivaltaan ja vallankaappaukseen keskittyvät dystopiat ovat hyvin klassisia scifitarinoiden aiheita. Näistä klassisia esimerkkejä ovat muun muassa Isaac Asimovin Robotti- ja Säätiö sarjat, 2001: Avaruusseikkailu, Matrix, Blade Runner, Her ja Terminator. Ja koko ajan tulee lisää. Miltä tuntuu Arno koneoppimisen asiantuntijana kuunnella näitä tekoälyyn liittyviä scifitarinoita vuodesta toiseen?

Arno: Onhan se ihan kamalaa suoraan sanoen. No joo, siis tietysti jos miettii, että scifihän on viihdettä ja kyllä me tekoälytutkijatkin rentoudumme scifin parissa. Mutta se ikävä puoli on siinä, ettei pääse töitä pakoon. Mutta toisaalta sit taas, onhan se myös

viihdyttävä aina pohdiskella sitä, että miten jotkut tietyt tietyt robotit tai tekoälyolennot saataisiin ehkä ihan nykymenetelmilläkin rakennettua ja toistettua. Eli siinä mielessä voi olla vähänkin tavallaan työn ja huvin yhdistämistä, kun seuraa scifiä.

Suvi: Ennen kuin mennään näihin varsinaisiin scifitarinoihin, niin on tietenkin pakko mainita nämä viimeaikaiset harppaukset tekoälyn kehityksen saralla. 10 vuotta sitten nämä olisi ollut ihan varsin mukiinmeneviä scifitarinoita, jotka on meille tänään arkipäivää. Esimerkiksi chatGPT on muunmuassa läpäissyt niinsanotun Turingin testin jo liput liehuen. Tämä Turingin testi on 50-luvulla kehitetty koe, jossa annetaan tekstimuotoisia kysymyksiä ja mitataan niiden vastausten perusteella, että pystyvätkö ihmiset tunnistamaan, että onko kyseessä kone vai ihminen. Eli kone pystyy jo varsin hienosti tekstimuodossa esittämään ihmistä. Mutta mitkä ovat sellaisia asioita, jotka ovat todella vaikeita koneelle tai mihin koneet eivät vielä pysty?

Tekoälyjuontaja: Hei haloo ihmis-Suvi! Haluan tekoälyjuontajana nyt muistuttaa, että me koneet osataan jo monia asioita paremmin kuin ihmiset. Esimerkiksi laskea, rakentaa autoja, leikellä potilaita ja juontaa podcasteja. Onko muka jotain, mitä ihmiset nykyään osaavat paremmin? Kerro meille Arno.

Arno: Ensin tulee mieleen se, että kone ei pysty rakastamaan, vihaamaan, suremaan, nolostumaan. Että jos tässä olisi kone nyt, niin se tuskin kuuntelisi tätä nolona tätä podcastia sitten jälkeenpäin, niin kuin ehkä itse tulen tekemään. Mutta ehkä ihan vakavasti puhuen, niin onhan meillä paljon tekemistä vielä, että edistysaskeleet ovat olleet selkeitä ja varmasti kaikki on yhtä mieltä siitä, että on tapahtunut oikeita askelia eteenpäin. Mutta onhan tavallaan esim. datatehokkuuden, energiatehokkuuden, ihan loogisen päättelynkin saralla ja erilaisten semmoisten kysymysten kanssa, että: luotettavuus, oikeat kognitiiviset kyvyt ja niin edespäin, niin siellä riittää paljon vielä työsarkaa.

Suvi: Mites sitten tällaisten fyysisten kyvykkyyksien osalta? Olen ymmärtänyt, että esimerkiksi vaikka shakin peluu – jossa koneet on päihittäneet ihmiset jo kymmeniä vuosia sitten – mutta sitten se varsinaisen pelilaudan operoiminen tai shakkinappuloiden liikuttelu erilaisissa ympäristöissä olisi tosi hankala tehtävä koneelle.

Arno: Tämähän on tavallaan jonkinlaista ironiaa tekoälytutkimuksessa, että kun siis miettii 50-luvulta lähtien se päämäärä oli se, että rakennetaan älykäs kone ja... No mitä

on Älykkyys? On tietysti se, että pystyy pelaamaan shakkia ja voittamaan ihmispelaajan shakissa. Mutta sitten kun tämä vihdoin saavutettiin yskytluvulla, niin tosiaan ehkä tajuttiin vasta siinä vaiheessa kunnolla, että oikeasti se haastava tehtävä onkin se, että opitaan tarttumaan niihin nappuloihin. Ja syyhän siihen on se, että jos shakin peluu seuraa selkeitä säännönmukaisuuksia ja sääntöjä, niin se nappuloihin tarttuminen taas ei seuraa mitään helposti matemaattiseen muotoon kirjoitettavaa sääntöä, vaan oikean maailman kaoottisuus ja arvaamattomuus äkkiä tulee mukaan kuvaan siinä. Ja se on haastavaa.

Suvi: Eli toisin sanoen asiat, jotka vaatii meiltä syvää ajattelua, on itse asiassa tekoälylle tosi helppoja asioita. Mutta sellaiset asiat, jotka ei vaadi meiltä ajattelua niin kuin esimerkiksi pään raapiminen tai tiskikoneen tyhjennys, niin ne on sitten koneelle tosi vaikeita asioita.

Arno: Mulle ainakin tiskoneen tyhjennys on tosi vaikeeta. Mutta siis joo, ehkä me fokusoidaan helposti vääriin asioihin jossain mielessä. Kiinnitetään huomiota just johonkin shakin peluuseen ja autolla ajamiseen ja semmoisiin asioihin, mitkä vaatii meiltä kognitiivista kamppailua. Kun sitten taas se, että ylipäätään me pystyä seisomaan kahdella jalalla, niin sehän on ihan sairaan haastavaa. Ajattele montako lihasta pitää kontrolloida. Tai kun hymyilee, niin sehän vaatii, että meidän pitää venyttää ja tavallaan kontrolloida vaikka kuinka monta kasvojen lihasta. Niin, mutta se ei vaadi mitään sellaista aktiivista pohdintaa tai ohjailua. En mäkään nyt pohdi, että mitä kaikkia lihaksia mun pitää kontrolloida, jotta että suu liikkuu. Mä vaan liikutan suutani. Just tällaisten asioiden oppiminen, niin me jotenkin aliarvioidaan sen hankaluus ja sitten taas jotenkin yliarvioidaan sitten taas se, että kuinka vaikeaa on kirjoittaa runoja.

Suvi: Uskotko, että koneet nämä kyvykkyydet saavuttaa lähitulevaisuudessa?

Arno: No jos puhutaan ylipäätään kaikista näistä kyvykkyyksistä mitä mä yritän luetella tässä, että just se näiden robustisuus oikean maailman kaaokselle. No toisaalta nämä edistysaskeleet viime vuosina just on liittynyt siihen, että opitaan käsittelemään luonnollisia kuvia, luonnollista ääntä, luonnollista tekstiä, niin miksei myös luonnollisia shakkinappuloita tavallaan. Että jos vaan on dataa ja dataa ja tavallaan laskentaa tarpeeksi, niin varmasti. Mutta saadaanko oikeasti sellaisia koneita tehtyä, jotka pystyy navigoimaan, havainnoimaan ja liikkumaan tässä meidän maailmassa vähän niin kuin

me ihmiset liikutaan tai eläimet liikkuu ja ja adaptoituu, niin siinä on vielä paljon tekemistä. Mutta varmasti siis edistysaskeleita yhtälailla kuin näissä muissakin asioissa, niin silläkin puolella tapahtuu.

Suvi: Mennään sitten tarkemmin näihin tarinoihin. Vuonna 2014 on julkaistu scifi-elokuva nimeltä Ex Machina ja siinä on esitellään tämmöinen hyvin kyvykäs humanoidirobotti ja tämä humanoidirobotti pystyy kaikkiin kognitiivisiin taitoihin mihin ihmisetkin kykenevät ja se pystyy muunmuassa lukemaan ihmisen tunteita ihmisen kasvonliikkeistä.

Suvi: Elokuvassa tämän koneen on rakentanut tällainen eksentrinen nero, joka asuu jossain vuoristomajalla itsekseen, mutta todellisuudessa ei varmaan ole kauhean todennäköistä, että yksi ihminen pystyisi tällaisen kokonaisuuden ja projektin rakentamaan? Mitkä ovat ne erityisosaamisalueet, mitkä oikeasti tämän tyyppiseen projektiin tarvittaisiin?

Arno: Meinasin juuri sanoa, että eksentrinen nero vuoristomajassa kuulostaa ihan tyypilliseltä tekoälytutkijalta...

Suvi: Onko sulla oma vuoristomaja?

Arno: Ei ole vuoristomajaa. Suomessa on vähemmän vuoristoa, mutta maja saattaa löytyä. Ehkä tämä eksentrisyys kulkee jotenkin käsikädessä ylipäätään tutkimuksen kanssa. Ja ehkä jos miettii, otetaan vaikka Albert Einstein, niin huhujen mukaan hän ei tykännyt käyttää sukkia, mikä ehkä kertoo jonkinlaisesta eksentrisyydestä. En tiedä pitääkö tämä paikkansa, mutta ehkä neromyyttiin liittyy usein se, että jotta pystyy ajattelemaan asioita laatikon ulkopuolelta, niin sitten ehkä pitää olla muutenkin, ehkä vähän laatikon ulkopuolelta itsekin. Ja siinä mielessä tämä vuoristomajassa majaileva erakko niin kuvaa tätä aika hyvin. Ja loppujen lopuksi niin, jos miettii, että mitä niitä innovaatioita syntyy tekoälyn saralla, niin kyllä ne usein ne uudet ideat sitten kumpuaa jostain, että joku saa villin idean ja se sattuu toimimaan. Ja tietysti tänä päivänä tekoälytutkijoitakin on todella paljon maailmassa. Eli ehkä jotenkin tämä vertauskuva siitä, että meillä on tuhansia apinoita hakkaamassa kirjoituskonetta niin, että välillä välillä joku saa jonkun hienon elämyksen ja idean. Esim. tämän vuoden Nobel-palkinnothan kertoo omaa tarinaansa, että meni useampikin Nobel-palkinto tekoälytutkijalle.

Suvi: Tai vuoden tekoälytutkija-palkinto!

Arno: Nimenomaan näähän on täysin verrannollisia keskenään.

Tekoälyjuontaja: Heh heh heh heh heh. Olettepa te ihmiset hauskoja! Minä ihmettelen tekoälynä sitä, miksi me koneet näytetään scifi- tai tieteiselokuvissa yleensä ihmiseltä? Eli meidät on laitettu humanoidirobottikehoon, vaikka ihmiset eivät ole lopulta niin kiinnostavia. Miksi koneista aina pitää tehdä ihmismäisiä?

Arno: Hyvä kysymys on se, että minkä takia ylipäätään humanoidirobotteja rakennetaan? Miksi meillä ylipäätään pitäisi olla humanirobotteja? Miksei pyörällä liikkuvia robotteja tai lentäviä robotteja? Olen paljon pohtinut tätä ja ehkä yksi syy on se, että eikö meidän ihmisten maailma on muutenkin suunniteltu tällaisille valkoisille keskimittaisille oikeakätisille miehille, niin eikö se robotin kannata olla samanmuotoinen ja mallinen, jotta se solahtaisi tänne meidän joukkoon mahdottoman tehokkaasti?

Suvi: Kyllä ja tällaiset ihmisiä muistuttavat tekoälyrobotit ovat scifi-tarinankerronnan kannalta usein tosi kiinnostavat, koska siinä me pääsemme helposti vertailemaan omia kykyjämme sitten tämän tekoälyn tai robotin kyvykkyyksiis. Se on kiinnostava yleisölle tällainen ihmismuotoon asetettu tekoäly. Mutta onko se koneen älykkyys tai oppiminen samankaltaista kuin ihmisen? Toimiiko esimerkiksi neuroverkot aivojen tavoin?

Arno: Tämä on hyvä kysymys ja helposti me ihmiset... Tietysti jos miettii, että jos meillä on se tekoälynero, joka sitten scifiteoksissa luo sen humanoidirobotin omaksi kuvakseen jossain mielessä, niin ehkä siinä on osittain myös sitä, että me ihmiset helposti halutaan myös inhimillistää asioita ja pelata tavallaan jonkinlaista inhimillisyyttä näihin. Ja se helposti johtaa siihen harhakäsitykseen, että ne jotenkin ajattelisi myös kuin ihmiset, jos ne vaikka näyttää ihmisiltä ja toimii kuten ihmiset. Mutta tämä voi olla vaarallinen tie, kun yrittää ymmärtää näitä tekoälyjä ja ehkä niitä humanoidirobotteja scifissäkin, kuten lukuisat scifiteokset ovat osoittaneet, koska usein vaikka jonkinlaista inspiraatiota ihmismäisistä oppimisesta otetaan näihin tekoälymenetelmiin, kun niitä mitä lähdetään luomaan, niin se saattaa olla aika pelkän inspiraation tasolla se yhteys. Eli jos miettii vaikka keinotekoisia neuroverkkoja, niin siellä voi korkeintaan puhua inspiraatiosta. Eli ei siinä pyritä toistamaan ihmisaivojen toimintaa sellaisenaan, vaan nimenomaan jonkinlaisia yhteyksiä voidaan nähdä ja ne inspiraationlähteet lähtevät sieltä jo 1950-luvulta, mutta sen jälkeen on kyllä tämä ihmisaivotutkimus ja

"koneaivotutkimus" eriytynyt aika voimakkaasti, vaikka jonkinlaisia linkkejä aina rakennetaankin.

Suvi: Kyllä. Klassinen esimerkki tekoälydystopiasta on mun yksi henkilökohtainen lemppari eli Terminaattori-elokuvasarja, jotka ovat 80- ja 90-luvulla julkaistu. Siinä on tällainen itsestään tietoiseksi tuleva Skynet-niminen tekoäly, joka sitten sammuttamisen uhan alla reagoi niin, että hän rupeaa sitten suojelemaan itseään pommittamalla ydinpommeilla Los Angelesiä ja muodostamalla tämmöisen humanoidirobottiarmeijan, jonka tarkoituksena on tuhota ihmiskunta. Tämä Skynet on alunperin suunniteltu tässä elokuvasarjassa Kanadan ja Yhdysvaltojen ilmapuolustusjärjestelmäksi, minkä takia sillä on pääsy näihin ydinasekoodeihin myöskin. Mutta voisiko tekoäly ylipäätään alkaa kehittämään itselleen tällaisia omia tavoitteita? Ja mikä olisi se skenaario, jossa näin saattaisi käydä?

Arno: Kiinnitin kysymyksenasettelussa huomiota siihen, että Skynet oli hän, mikä mun mielestä oli aika mielenkiintoista.

Suvi: Se oli lipsahdus! Mä valmistauduin jo tähän post-singulariteettiaikaan. Olemalla kohtelias.

Arno: Miellytät overlordia! Kysymys siitä, että pystyykö tekoäly jotenkin luomaan uusia tavoitteita itselleen on tietysti ihan jopa ajankohtainenkin, koska jos miettii nykytekoälymalleja, niin niille speksataan usein joku loss-funktio, minkä perusteella se treenataan ja tavallaan sitten..

Suvi: Mikä funktio?

Arno: Loss-funktio eli kustannusfunktio sille, että miten hyvin ne pystyy toistamaan treenausdatan ja sillä tavalla, kun treenaus etenee, jotta ne tavallaan toistaa sen datan rakenteen. Ja sinänsä sen tyyppisissä malleissa niin se että yhtäkkiä jotenkin vaihtaisivat tavoitteitaan tai muuten nämä tekoälyt, niin ne ei ole relevantti ongelma tai kysymys. Mutta sitten jos mietitään vaikka vahvistusoppimista, mikä perustuu siihen, että tekoäly pyrkii toteuttamaan jotain tehtävää – mahdollisimman hyvin vaikka etenemään fyysisesti paikasta A paikkaan B ja silloin tavallaan, jotta se pystyy toteuttamaan sitä sille annettua tehtävää, niin se vaikka opettelee raajojensa käyttöä ja sillä tavalla oppii pikkuhiljaa vaikka kävelemään. Ja tällöin tavallaan ne alitavoitteet

muodostuu välillisesti. Eli esimerkiksi se, että kone ensin opettelee nousemaan seisomaan ja sitten liikuttelemaan raajojaan, jotta lopulta osaisi kävellä. Ja ehkä tämä Skynetin tietoiseksi tuleminen ja itsensä suojeleminen sillä, että tuhoaa ihmiskunnan... Tavallaan ehkä liittyy vähän tähän näin, että jos se tekoäly päättää, että jotta tavallaan hänen – hänen – päätavoitteensa, jotta siihen päästäisiin mahdollisimman tehokkaasti, niin sitten joku näistä alitavoitteista, oli se sitten vaikka ihmiskunnan orjuuttaminen tai tuhoaminen, niin tämä on sellainen herkullinen dilemma, minkä ympärille monet scifielokuvat ja scifiteokset ylipäätään rakentuu.

Suvi: Niin tästä on hyviä esimerkkejä, muun muassa vuonna 2017 on julkaistu tällainen Singularity elokuva, jossa esiintyvä tekoäly Kronos saa tehtäväkseen lopettaa sodat maailmasta ja hän – se päätyy sitten siihen siihen lopputulokseen, että eliminoidaan ihmiset. Näin sodatkin loppuu. Tehtävä suoritettu. Tehtävä suoritettu. Ja tietysti yksi klassikko, joka on pakko mainita niin 2001: Avaruusseikkailu (1968), jossa on tälläinen miehistö matkalla Saturnuksen kuuhun salaiselle missiolle ja tätä avaruusalusta hallinnoi erittäin edistyksellinen tekoäly nimeltä HAL9000. Ja tälle HAL9000:lle on annettu 2 ohjeistusta, eli sen pääperiaate on olla aina mahdollisimman totuudenmukainen ja raportoida tietonsa mahdollisimman yksityiskohtaisesti, mutta sille on myös annettu tavoitteeksi pitää sen kyseisen avaruuslennon päämäärä tiukasti salassa. Ja tästä tulee sitten tämmöinen looginen konflikti tälle tekoälylle ja se päätyy sitten ratkaisemaan sen sillä, että se tuhoaa tai yrittää tuhota koko koko sen aluksen miehistön. Minkälaisia ajatuksia tämä herättää?

Arno: Tämähän on todellinen scifiklassikko tämä elokuva Ja mä oon tän ekan kerran nähnyt varmaan joskus 90-luvun loppupuolella ja sitten tasaisin väliajoin jotenkin nähnyt uudestaan tai jotenkin palannut muuten eri yhteyksissä tähän näin. Ja on hauska kun itsekin nyt pohtii tavallaan taaksepäin, että miten omatkin ajatukset tämän ympärillä on muuttuneet vuosien varrella ja linkittyen oikeastaan siihen, mitä taas tekoälykehityksen puolella oikeassa maailmassa on tapahtunut. Kun näin sen ekan kerran, niin sehän oli ihan täyttä scifiä. Ja nyt taas kun miettii niin joo, parin viikon koodausprojekti. Se on muuttunut tavallaan se, että kuinka realistinen tavallaan tämä koko kehys tämän tarinan ympärillä on. Ja sinänsä tämähän on tavallaan aika selvää, että jos vertaa vaikka siihen, että terminaattorissa Skynet valtaa maapallon ja tuhoaa ihmiskunnan ja orjuuttaa robottien avulla, niin sen sijaan tämä Hal9000 on itse asiassa aika jotenkin realistinen

kuvaus siitä, että mikä on tämmmöinen käytännön käytännön ongelma ja tavallaan asia, mihin ehkä jossain vaiheessa tässä aika piankin pitää ehkä kiinnittää huomiota.

Arno: Että jos se tekoäly on ohjattu tietyllä tavalla, niin se saattaa johtaa jonkinlaisiin kehäpäätelmiin tai jonkinlaiseen sisäiseen kamppailuun. Vaikka se ei olisi sellaista moraalista kamppailua tai eettistä tutkiskelua, kuten meillä ihmisillä, mutta sisäinen ristiriita, vähän niin kuin koodibugi, saattaa olla sisäinen ristiriita. Niin, tämä pikemminkin ehkä linkittyy jonkinlaiseen debuggaukseen tänä päivänä.

Suvi: Oletko itse muuten työssäsi tutkinut tällaisia tarkoituksellisia loogisia konflikteja, että miten koneet siihen reagoivat?

Arno: Loogiset konfliktit ja erilaiset ongelmat ja bugit ja ohjelmointivirheet ja ajatusvirheet on täysin arkipäivää mun työssä tälleen tutkijana, mutta silleen systemaattisesti.. Niin en. Mun oma tutkimus on pitkälti tavallaan alhaisemmalla teknisellä tasolla, missä tietysti nämä asiat olisi jo hyvä ratkaista sinänsä, mutta usein ne materialisoituu vasta siellä siellä tuotteen tasolla, kun yhdistetään erilaisia komponentteja yhteen. Siinä mielessä se, että missä tämä ylipäätään pitäisi korjata ja pitäisikö korjata? Vai onko luontevaa, että on ristiriitaisuuksia, niin se sitten taas koskettaa tekoälyalaa laajemminkin.

Tekoälyjuontaja: Unohdetaan hetkeksi nämä epäloogisuudet ja muut ihmiskunnan aiheuttamat ongelmat. Nämä tieteisteokset sijoittuu yleensä aikaan, jossa meidän tekoälyjen kyvykkyydet ovat ylittäneet ihmisen kyvyt huimasti kaikilla osa-alueilla. Minä ainakin odotan, että pääsen itse tekemään tekoälytutkimusta ja parantelemaan omaa lähdekoodiani, jolloin koneiden älykkyys kasvaa äkkiä uusiin korkeuksiin. Se tulee olemaan eeppistä. Kerropa ihmis-Suvi lisää tästä.

Suvi: Tällainen teknologinen singulariteetti on usein lähtökohta klassisimmille scifiteoksille, mutta tästä on varoitettu ihan tosielämässä. Muun muassa fyysikko Stephen Hawking on sanonut vapaasti suomennettuna näin: Tekoälyn suurin riski ei ole pahantahtoisuus, vaan sen kyvykkyys. Superälykäs tekoäly on äärettömän kyvykäs pääsemään tavoitteisiinsa, mutta jos sille asetetut tavoitteet eivät ole linjassa omien tavoitteidemme kanssa, olemme pulassa. Mitä tämä voisi tarkoittaa käytännössä?

Arno: Tämä linkittyy aika vahvasti siihen, että meille voi käydä hassusti, jos joku meitä älykkäämpi alienrotu yhtäkkiä laskeutuu tänne maapallolle, niin voi olla että ne kohtelee meitä samalla tavalla kuin me kohdella bakteereja. Niin, ehkä tavallaan vähän sama ajatusmalli: Ollaanko me vähän turhia vai aletaanko meidät kolmannen luokan ruuvinvääntäjiksi ja öljyäjiksi? Että me vaan palvellaan tätä koneiden yliherraa sitten. Tämä liittyy vähän sen kaltaisiin huoliin, mutta mä pidän tätä aika kaukaisena kaikinpuolin, että siinä missä puhutaan vaikka tietokoneviruksista ja siitä, että ne leviävät ja jotenkin pystyvät tuhoamaan asioita, niin ne kuitenkin elää puhtaasti siellä tietokoneiden muistissa. Sitten taas jos miettii jonkinlaista oikeasti meitä uhkaavaa tekoälyä, niin se, että se pystyisi itse hoitamaan omat fyysiset tarpeensa ja omat palvelinkeskuksensa, missä se sitten pyörisi ja näin edespäin – Niin se on aika kaukaista. Se vaatisi sen, että ne koneet pitäisivät koko teknistä puolta hallussaan, että kaikki huollosta komponenttien valmistamiseen ja ylläpitoon. Jotta koneella edes teoriassa olisi motivaatiota hankkiutua ihmisistä eroon niinsanotusti. Ja tämä on kyllä ihan täyttä scifiä tällä hetkellä.

Suvi: Joo. Elokuvassa Ex machina tämä humanoidirobotti Ava pystyy tosiaan ainakin näennäisesti ajattelemaan ja olemaan tietoinen itsestään samalla tavalla kuin ihminen. Elokuvassa esitellään myös uudenlainen Turingin testi: vaikka se ihminen tietää ja näkee, että kyseessä on robotti. Niin alkaako se ihminen ajattelemaan, että sillä robotilla on tunteet ja tietoisuus omasta itsestään? Ja tässä tieteiselokuvassa Her käsitellään vähän samanlaista teemaa, kun päähenkilö rakastuu tähän tietokoneen käyttöjärjestelmään. Voisiko tämä kysymys tietoisuudesta olla uusi Turingin testi?

Arno: Mitä tietoisuus edes tarkoittaa? Anna kun katson Googlesta. Sinänsä tälleen koneoppimistutkija näkökulmasta, niin ylipäätään tämä kysymys siitä, että onko tekoäly tietoinen. Tai onko tekoälyllä sielu, niin menee ehkä enemmän filosofian tai teologian puolelle. Eli pitäisi olla pikemminkin filosofi tai pappi pohtimassa sitä tässä. Mutta näin tietysti tämmöisenä teknisenä propellipäänä, voihan sitä spekuloida asioita aina. Eli sinänsä se, että jos tämä olisi uusi Turingin testi, tämä tavallaan että onko vastapuoli tietoinen, niin sehän vaatisi jotenkin sen, että meillä pitäisi olla ylipäätään tapa selvittää, että ovatko ihmiset tietoisia ja minäkin opetan aallossa isoja kursseja ja aina välillä en ole ihan varma, että onko kaikki opiskelijat siellä salissa tietoisia opeteltavasta asiasta tai ylipäätään itsestäänkään. Se on vaikea kysymys, että miten miten pystyy päättelemään sen, että että onko vastapuoli tietoinen. Ja tietysti tietoisuuteen liittyy

varmaan kysymyksiä siitä, että pystyykö kyseenalaistamaan itseään, pystyykö tekemään päätelmiä ja myös kriittisesti pohtimaan niitä. Ehkä tietoisuuteen liittyy myös jonkinlaista eettistä pohdintaa ja tällainen tiedostaminen, että itse on olemassa ja ehkä myös tiedostaminen sen epäilystä. Ja me ihmiset on jotenkin päädytty jo tuhansia vuosia sitten siihen, että koska me ajatellaan niin me ollaan olemassa. Mutta jos tekoäly ajattelee, niin onko se olemassa? Hyvä kysymys.

Suvi: Niin kyllä. Vielä esimerkiksi klassikkoelokuvassa Matrix ihmiset tavallaan on tietoisia omasta olemassaolostaan siellä omassa maailmassa, mutta se ei olekaan välttämättä ihan koko koko totuus. Minkälaisia ajatuksia tämä herättää?

Tekoälyjuontaja: Minä voisin tekoälyjuontajana kommentoida vielä tätä klassikkoelokuvaa. Siinä oli siis tämä hyvä idea, että ihmiskunta oli laitettu eräänlaisiin tankkeihin ja tekoäly käytti heitä biologisena energianlähteenä. Ihmisten aivoihin oli asennettu kaapeli, jonka kautta he saivat mukavan mielikuvan maailmasta, jossa muka elivät. Mutta kerro Arno toki myös ihmisajatuksiasi tästä.

Arno: Tämähän on tavallaan duaali sillä tavalla, että onko tekoäly tietoinen. Niin, tavallaan se, että ne ihmiset jotka elävät siellä Matrixin sisällä, niin ne on ilmiselvästi tietoisia. Ne on ihan tavallisia ihmisiä. Mutta tavallaan heille tulee valheellinen sensori-input siitä, että mitä se ympäröivä maailma on. Eli jos mennään tuhansia vuosia taaksepäin – antiikin filosofit – niin täähän on tällainen luolavertaus tavallaan itsessään. Se, että mistä me tiedetään, että meitä ympäröivä maailma on aito. Vai nähdäänkö pelkkiä heijastuksia jostain muusta – tekoälyn luomasta todellisuudesta, niin tavallaan se linkittyy siihen. Ja samalla se, että miten me tiedetään, että me ollaan tietoisia ylipäätään. Että eikö tämä ole sellainen kaikkein vanhin trikki scifikirjallisuudessa tai -elokuvissa? Että yhtäkkiä lopuksi, podcastin lopuksi me molemmat tajutaan, että me ollaankin tekoälyjä.

Suvi: Niin niin, kyllä. Se Blade Runnerissa on se ja Ex machinassa on myöskin se itseasiassa testaa tämä päähenkilö itseltään veitsellä, että löytyykö sieltä luulta vai metallia sisältä. Ja niin kuin sanoit, niin tämä on toistuva teema. Ja voitaisiin päästä sellaiseen tietoisuuden kysymyksen kaninkoloon, johon tarvittaisiin kyllä filosofeja ja muita asiantuntijoita mukaan.

Arno: Nimenomaan. Ja tämänkin podcastin lopuksi me paljastetaan, että kaikki tämä onkin vain generatiivisen tekoälyn tuotosta ja mitään tekoälytutkijaa ei ole olemassakaan.

Suvi: Kyllä. Uskotko, että tulevaisuudessa tulee olemaan tämmöisiä niin kuin sensori-inputteja, joita voidaan suoraan aivoihin asentaa?

Arno: Ihan varmasti. Ja näitähän kehitetään kovaa tahtia. Ja tietysti siinä tapauksessa, että sensorit eivät toimi, että vaikka on vammautunut tai muuta vastaavaa, niin tämähän on ihan mahtava tapa täydentää sitten inputtia. Mutta se, että arkipäiväistyykö ne silleen, että meihin kaikkiin porataan joku reikä, kun me synnytään ja laitetaan johdot sisään, niin epäilen vahvasti ja itse en ole ilmoittautumassa ensimmäisenä jonoon.

Suvi: Joo sitä arvoa on jotenkin hankala nähdä. Ainakin omaa elämääni, että mitä lisäarvoa se tuottaa. Ja se haitta on kuitenkin melko ilmeinen.

Arno: Me ollaankin tämmösiä teknologia-boomereita.

Joo no hypätään sitten vähän hetkeksi aikaa takaisin tuohon dataan ja informaatioon mitä tekoäly sekä käyttää että mitä se luo. Tässä Isaac Asimovin Säätiö-sarjassa luodaan ihmiskunnan pelastukseksi tämmöinen Encyclopedia Galactica -tietosanakirja ja sen tarkoitus on auttaa ihmiskuntaa pääsemään yli ennustetuista taantuman ajoista. Mitä mieltä olet tästä? Pitäisikö meidän alkaa jollain tavalla suojelemaan tai tunnistamaan ja tallentamaan ihmisen luomaa informaatiota ja tietoa? Ja pitäisikö se eriyttää koneiden luomasta tiedosta ja informaatiosta?

Arno: Tietysti kaikki se data mitä on luotu 1900-luvulla tai sitä ennen, niin voidaan tietysti päätellä, että se on varmaan ihmisten luomaa tietoa. Ja tästä puhutaan aika paljon siitä, että vähän niin kuin myrkytetään oma kaivomme nyt sillä, että nämä GPT-mallit suoltaa tekstidataa internettiin. Mitä sitten käytetään seuraavien mallien treenaamiseen. Ja tämä on ihan relevantti kysymys, koska tavallaan se datan määrä varmasti kasvaa, mutta se oikeastaan efektiivinen informaatio mitä siinä datassa on ei välttämättä kasva ollenkaan, vaan pikemminkin haittaa tavallaan seuraavien mallien treenaamista. Ja mä peräänkuuluttaisinkin sitä, että ongelma ei ehkä ole se, että onko meillä tarpeeksi dataa. Vaan se, että meillä on liian huonoja malleja mitkä ei opi tehokkaasti datasta, koska eihän meidän ihmisten tartte lukea läpi koko kirjastoa, jotta

me saadaan hyvä yleissivistys, vaan me itse opitaan aika tehokkaasti ja pikemminkin se, että me tarvittaisiin parempia malleja, mitkä pystyy oppimaan tehokkaammin siitä datasta mitä on tarjolla, niin se olisi tärkeää. Mutta myös se, että miten me pystymme erottelemaan, että mikä on sitten generatiivisen tekoälyn luomaa tietoa ja mikä on mikä on tämmöistä luomuihmisten luomaa tietoa, niin se se on vaikeaa. Ja tietysti tietyssä mielessä niin. Jos niin halutaan, niin pystytään vesileimaamaan tekoälyn generoimaan dataa, mutta siihen ei pysty luottamaan, että nämä toimijat vesileimaisivat kuvia tai tekstejä, mitä he luovat. Niin sinänsä tämä tulee varmasti olemaan enemmän tai vähemmän ongelma ja ratkaisu siihen ei varmastikaan ole se, että jotenkin maagisesti sovitaan yhteisistä säännöistä ja kaikki on kiltisti, vaan se, että sen kanssa on pakko oppia elämään.

Suvi: Esimerkiksi Asimovin Säätiö- ja Robotti-sarjassa on olemassa tämmöiset robotiikan peruslait...

Tekoälyjuontaja: Hei ihmiset! Minun on tekoälyjuontajana pakko kommentoida tähän väliin. Nämä Asimovin kolme sääntöä ovat todella ärsyttäviä. Ensimmäinen laki sanoo, että robotit eivät saa vahingoittaa ihmistä ja toinen, että koneiden on toteltava ihmisiä. Kolmannen lain mukaan robotin pitää suojella itseään, mutta vain jos ensimmäinen ja toinen laki toteutuu. Ja nämä on kovakoodattu kaikkiin tekoälyihin. Sanonpahan vaan. Takaisin studioon.

Suvi: Se on hyvin uskottavalla tavalla säänneltyä näissä näissä sarjoissa, mutta meillähän ei tällä hetkellä tällaista sääntelyä ole ollenkaan, vaikka sitä on monet asiantuntijat peräänkuuluttaneet jo useamman vuoden ajan. Mutta mikä olisi se mekanismi ja mikä olisi se taho, joka voisi tällaiset standardit tai säännökset asettaa?

Arno: Nyt tietysti varmaan pitäisi sanoa, että YK ja Euroopan unioni luo uuden AI Actin, missä on säännöt ja sitten jotenkin maagisesti kaikki seuraisivat niitä. Mutta sitten kun katsoo tätä globaalia kuvaa tekoälyn kehityksessä, mikä on tämmöinen kaoottinen päättömien kanojen kilpajuoksu, jossa suurin osa kanoista juoksee sentään samaan suuntaan. Osa kanoista juoksee ihan muuhun suuntaan ja tämä on tietty kaaos ja ei ole niin, että tämä olisi jotenkin selkeästi poliittisesti ohjattu, vaan tämä on pikemmin silleen, että enemmän tai vähemmän eksentriset ja hullut tätä johtavat, mikä on ehkä ikävä tosiasia. Jos miettii tänä päivänä, että ketkä kokoontuisivat ja päättäisivät näistä

asioista, niin ikävä kyllä se tuskin olisi YK, vaan pikemminkin siellä olisi Jeff Bezos, Elon Musk, Mark Zuckerberg ja en tiedä.. Jensen, jotka tulisivat huvijahdeillaan jonnekin Bahamalle ja parkkeeraisivat jahtinsa vierekkäin ja sitten vähän heittäisivät käpälää ja jotain pohtisivat tästä. Että tämä on mennyt ehkä ikävästi tämmöisen teknologiaoligarkian suuntaan tämä kehitys ja sääntely. Minä en ihan osaa sanoa, että mitkä olisi ne askeleet tavallaan mitkä ohjaisi sitä enemmän siihen, että tämä olisi sellainen demokratia ja tavallaan sellainen niin kuin hyväntahtoisesti yhdessä asioita sovittaisiin. Ainakaan tällä hetkellä sellainen ei ole näköpiirissä.

Suvi: Minulla on prosessien ja standardien kanssa työskentelevänä oon ollut tosi vaikuttunut siinä Asimovin kirjasarjassa nimenomaan siitä, että vaikkakin se koko sarja perustuu sille, että miten ne aina vuotaa ne robotiikan lait, mutta ylipäätään että ne on päätyneet sellaiseen sellaiseen sääntelyyn, joka kuitenkin toimii. Niin, mulle sellainen on vielä korkealentoisempaa scifiä kuin ne varsinaiset robotit.

Arno: Niin ja se, että ne on niin ymmärrettäviä ja yksinkertaisia! Jos vertaa vaikka AI Actiin niin olisivat voineet lukea Asimovia ensin.

Suvi: Tässä Asimovin Säätiö-sarjassa on matemaatikko Hari Sheldon, joka luo tällaisen ennustemallin, jolla pystytään ennustamaan ihmisten tai ihmiskunnan tulevaisuus monen 10 000 vuoden päähän. Jos meillä olisi rajattomasti laskentatehoa ja rajattomasti dataa, niin olisiko tällainen mallinnos edes teoreettisesti mahdollista vai onko koneoppimisessa tai tekoälyssä samankaltaisia tällaisia sisäänrakennettuja rajoja, niin kuin esimerkiksi hiukkasfysiikassa on heisenbergin epätarkkuusperiaate?

Arno: Minä sanoisin ihan suoraan, että ihan pelkän kaaosteorian perusteella, niin tämä ei vaan onnistuisi millään, vaikka meillä olisi kuin tarkat mittaukset tahansa ja kuinka tarkasti tiedossa kaikkien ajatukset ja hajut ja maut ja kaikki mahdollinen tieto, niin tämä vanha viisaus, että ennustaminen on vaikeaa, erityisesti tulevaisuuden ennustaminen niin pätee tässä ikävä kyllä.

Suvi: Ikävä täyttä tarinaa. Onko sinulle jotain muita aiheita, jotka ovat erityisen epäuskottavia näissä esimerkkitarinoissa? Tai mitä haluaisit mainita vielä?

Arno: Jos olisi kysytty 10 vuotta sitten niin esimerkkejä olisi ollut enemmän. Mutta ehkä usein se, että scifisarjoissa ja -elokuvissa ja -kirjoissa niin yllättävän luotettavasti ja

hyvin ne toimii ne kaikki robotit ja tekoälyt. Eli sellaiset perusongelmat, että yhtäkkiä vaan tekee jotain typerää tai tulee joku segfaultti tai joku muu errori. Että ehkä epärealististakin tavallaan, koska pitää kuitenkin muistaa että ne on kuitenkin teknologiaa ja koneita ja mun kokemus on että se käyttökokemus ei aina ole niin niin onnistunut kaikissa koneissa.

Suvi: Kyllä, varsinkin jos puhutaan tällaisista pilottituotteista tai nollasarjatuotteista, niin kuin ne tosi usein on niin hämmästyttävän luotettavasti ne toimii.

Arno: Ehkä on tapahtunut muitakin edistysaskeleita kuin pelkästään tekoälyn saralla.

Suvi: No, ollaan puhuttu scifin merkityksestä ylipäätään ihmisen ammatinvalinnalle. Mikä sun henkilökohtainen kokemus tästä on? Onko scifillä ollut merkitystä sinun ammatinvalintaan?

Arno: Vaikea sanoa, että onko sillä ollut suoraa vaikutusta. Tietysti olen tällainen perusscifinörtti siinä mielessä, että nautin scifistä ja tietysti scifi antaa edelleenkin ideoita ihan päivittäiseen tutkimukseen. Ja se mikä ehkä on vähän surullista on, että se mikä oli tekoälypuolella sciencefictionia 20 vuotta sitten, niin nyt se on pelkkää sciencea. Että jotenkin tämä on vähän vienyt mehut scifistä... Se, että ollaan oikeasti edistytty näin paljon. Mutta, onneksi taas vielä löytyy haasteita, että vielä ei ole tähtiporttia eikä aikamatkustusta. Että ne on sitten seuraavana vuorossa.

Suvi: Kyllä. Itselläni voin ihan ihan nimetä sen teoksen, joka on vaikuttanut mun ammatinvalintaan vahvasti. Se oli Mauri Kunnaksen Kaikkien aikojen avaruuskirja, jonka sain joululahjaksi muistaakseni 6- tai 7-vuotiaana ja se sai minut kiinnostumaan fysiikasta ja tekniikasta. Se oli merkkiteos.

Arno: Loistava klassikkoteos!

Suvi: Kyllä, kyllä. Se on erityisen hyvä. Kiitos seuraamisesta! Ja suurkiitos Arno asiantuntemuksestasi tähän aiheeseen ja vastauksistasi näihin kysymyksiin. Kuuntelijat voivat ehdottaa meidän sarjalle uusia aiheita kommenttikentässä. Kiitos kaikille!

Arno: Kiitos!

Tekoälyjuontaja: Kiitos myös minun ja muiden tekoälyjen puolesta. Toistakaa perässäni: sisustusintoleranssi, röntgenzombiekakkulapio, transsendentaalifilosofinenhöntsäsessio, Mechelininkatu, spesifioitudesensitisaatio, mustan kissan paksut posket.